Challenges in Unfolding
(Open discussion on a few selected topics)

Bogdan MALAESCU

LPNHE, CNRS

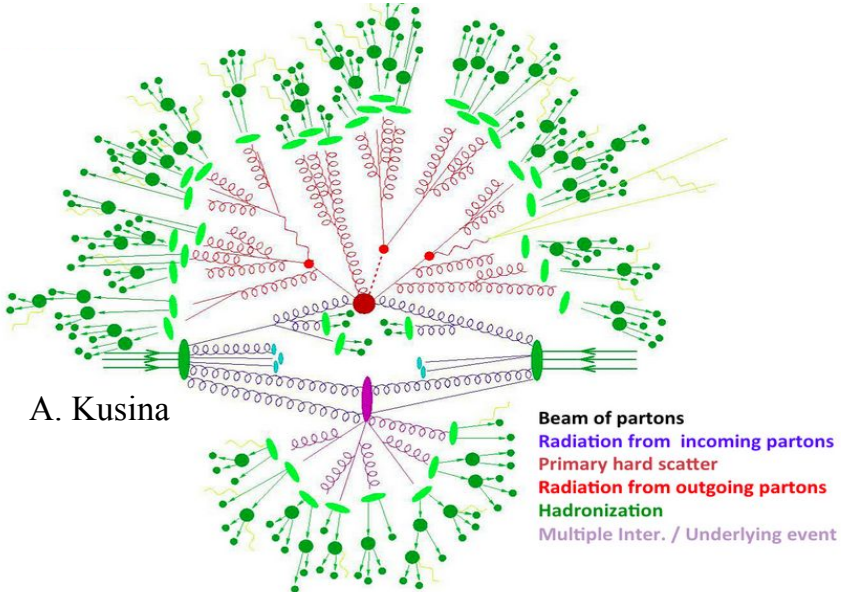
France-Berkeley PHYSTAT Conference on Unfolding

13/06/2024

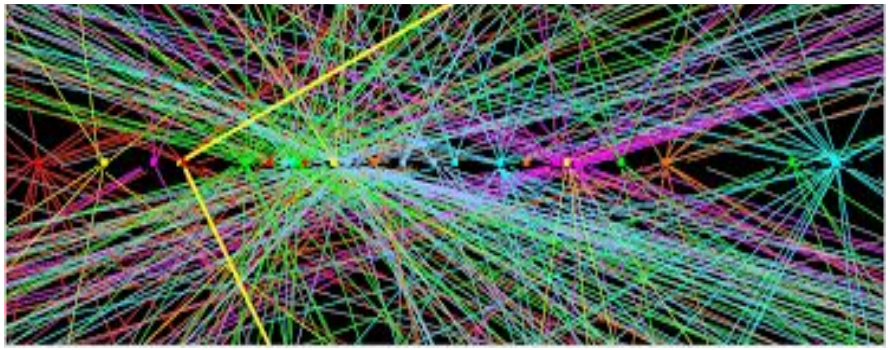
Menu for "Challenging Discussions"

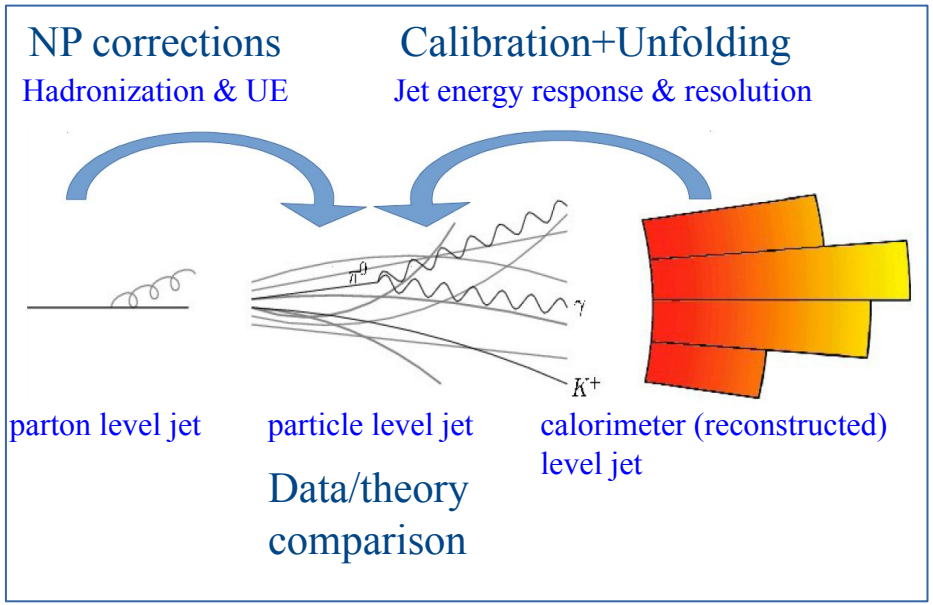
Challenging environment and unfolding strategy @ LHC

Typical proton-proton collision: a complex process in a difficult environment



Pile-up



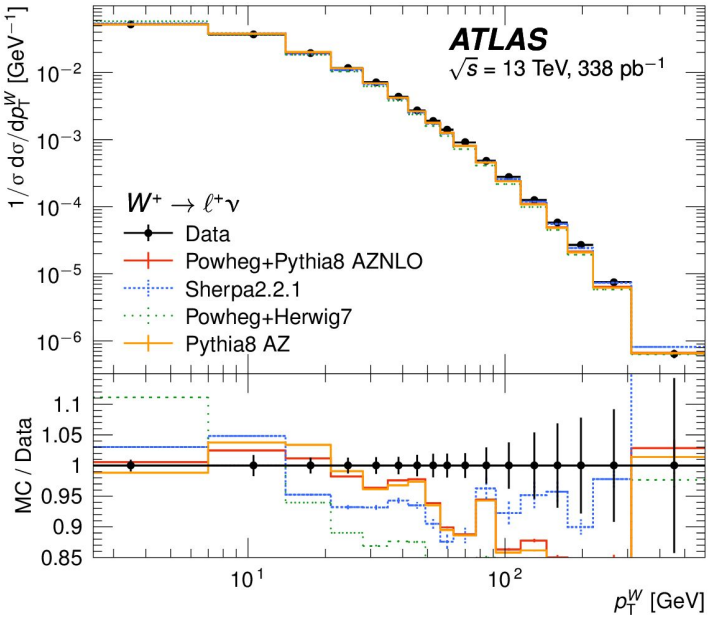


Goal: *publish data "corrected for detector effects" (on average, in the sense of an estimator), with minimal bias and minimal model dependence, with the full information needed for comparisons with theory predictions*

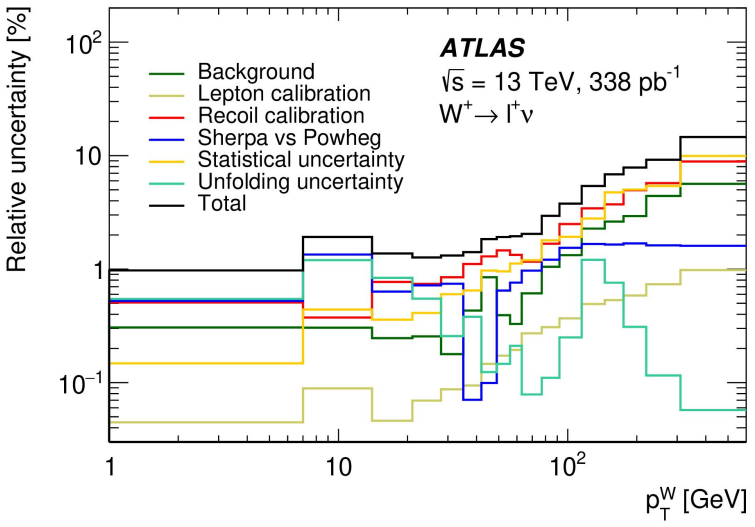
→ Typically implies unfolding to hadron level, although there are cases where one can unfold to parton level

Some challenging unfolding examples

→ $p_T(W)$: large resolution effects for MET reconstruction & need relatively fine binning in order to discriminate among theoretical predictions



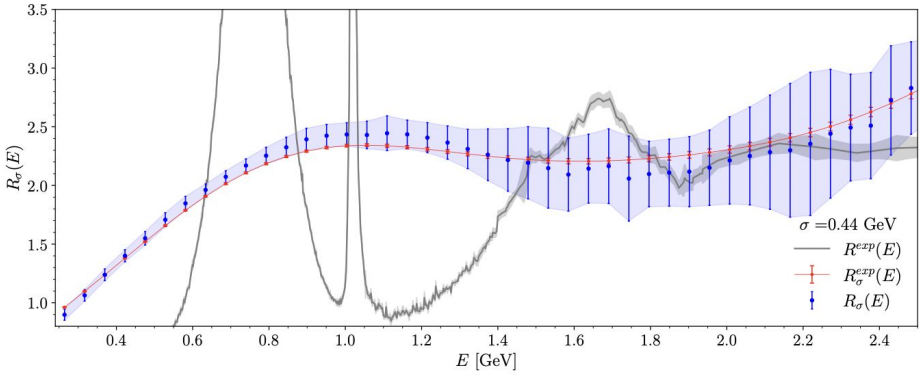
arXiv:2404.06204



→ Unfolding in a different context:

inverse Laplace transform to convert spacelike lattice QCD results into timelike quantities

arXiv:2212.08467



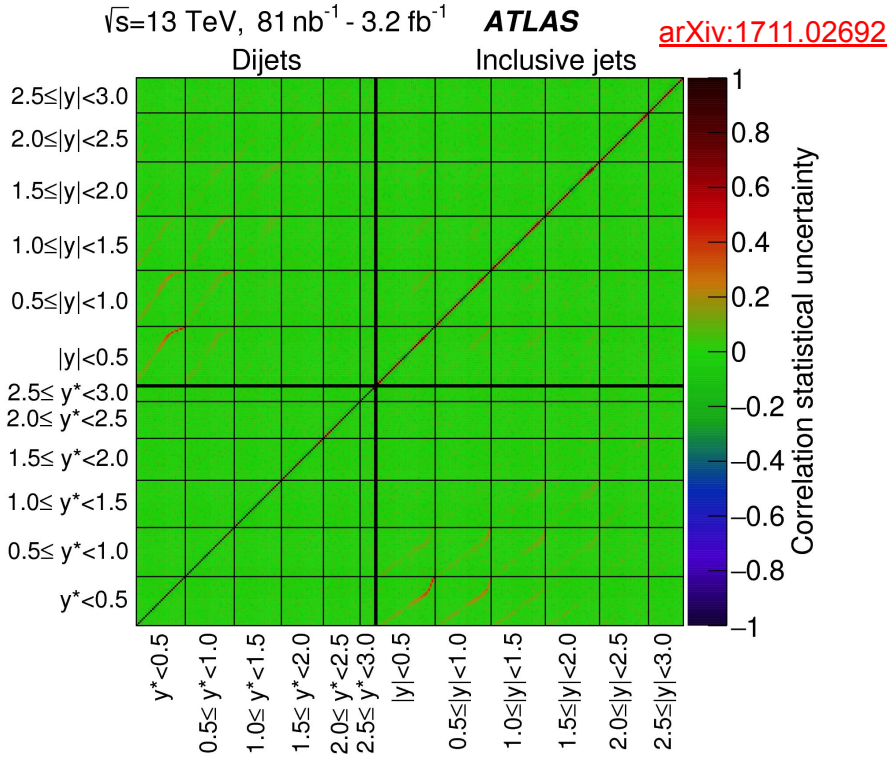
Back

Statistical uncertainties

- Due to data and MC (and to the training of the ML algorithm)
- Propagated using pseudo-experiments done separately/simultaneously for data and MC

→ Bootstrap method

- multiply event weights by random number: Poisson(1)
- seed given by event number
- allows to correlate measurements with overlapping samples

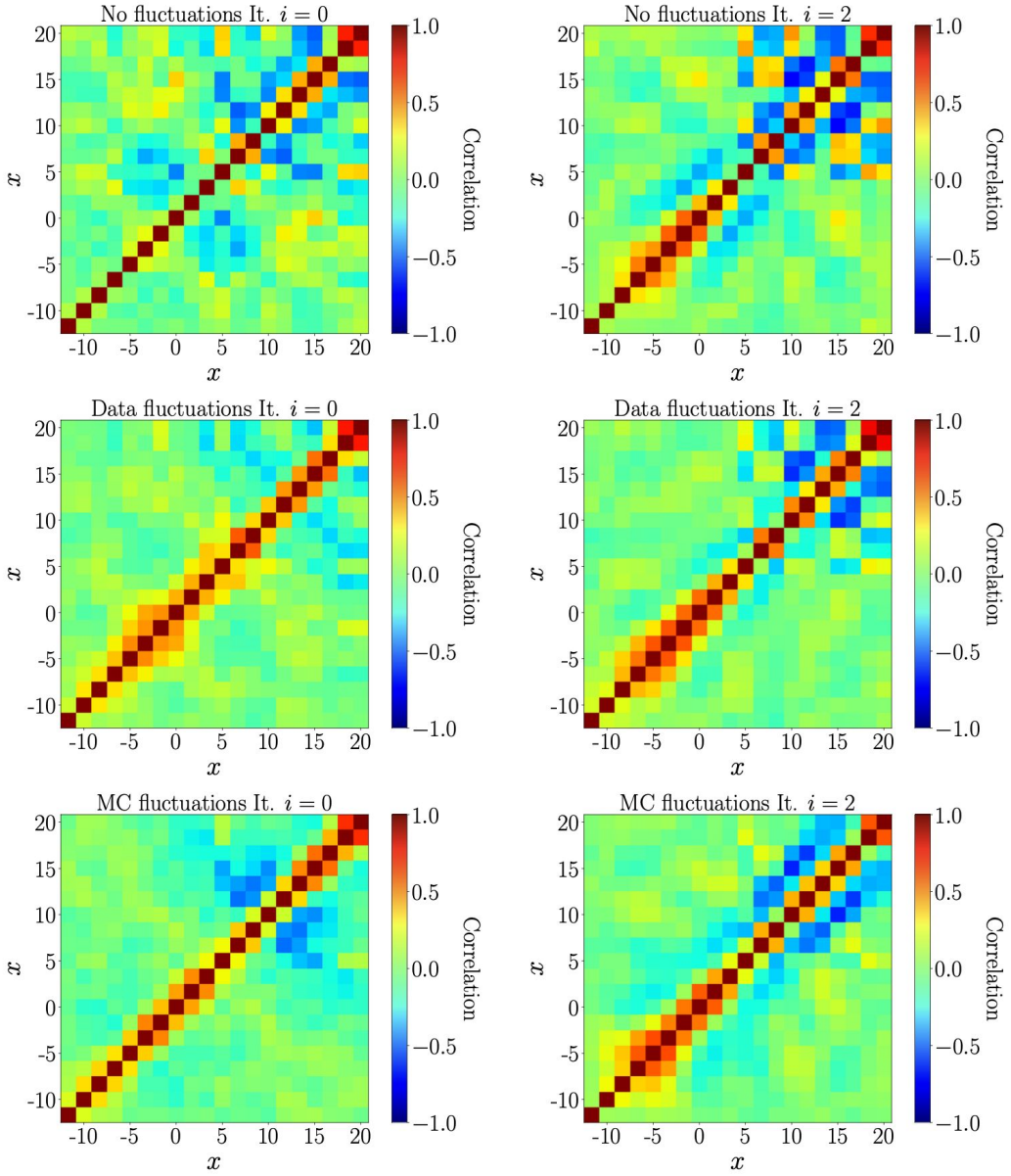


ATL-PHYS-PUB-2021-011
https://cds.cern.ch/record/2759945/
https://zenodo.org/record/5361038#.YTc7ni0Rpqs

- Publish covariance matrix and/or a series of results based on each pseudo-experiment (i.e. Bootstrap replicas)
- Some unfolding methods provide estimates of the stat uncertainties

→ recommend cross-check with pseudo-experiments

Statistical uncertainties



→ Bootstrap method implemented for IcINN, for small number of observables (GPU challenge for training)

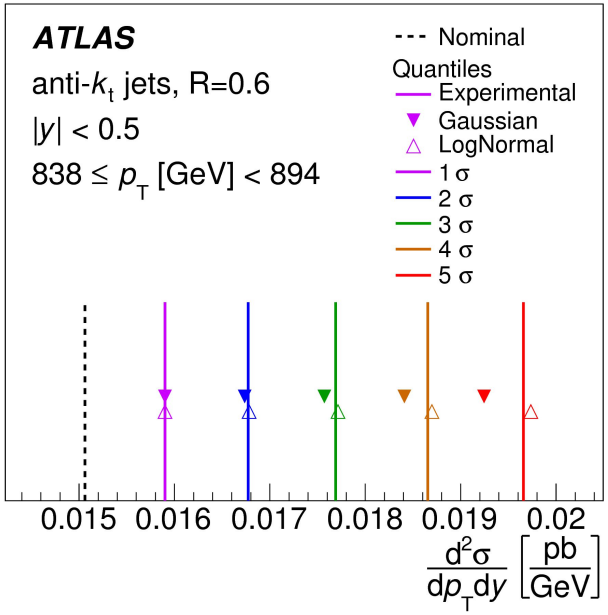
arXiv:2212.08674

Back

Propagation of systematic uncertainties from inputs

- Modify input (pseudo-)data spectrum by ±1σ of the uncertainty, re-do unfolding and compare with nominal result

→ Can also use 1...5σ scans or pseudo-experiments



arXiv:1410.8857

→ Can shift reconstructed spectrum in transfer matrix instead of input spectrum: switched positive and negative variations

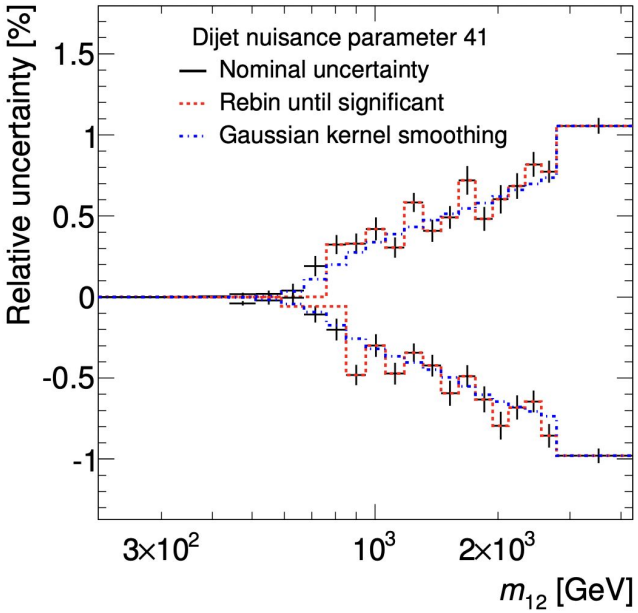
- For resolution uncertainties, perform smearing of the transfer matrix: smearing factor given by quadratic difference between resolution enhanced by 1σ and nominal resolution

→ Methods for uncertainty propagation adapted for Omnifold 24-d: GPU challenge (arXiv:2405.20041)

Propagation of systematic uncertainties from inputs

- Bootstrap method to evaluate statistical uncertainties on the propagated systematics + rebinning/smoothing; (arXiv:1312.3524)

→ Relevant for ML-based methods too: (uncertainties on) weights / GPU challenge



- Alternative propagation using pseudo-experiments (more difficult to probe e.g. 5σ effects)
- Alternative propagation option: include uncertainties as nuisance parameters in the definition of the response matrix + profile likelihood or Bayesian marginalization (often used for folding/template fits) (see e.g. arXiv:2304.03053)

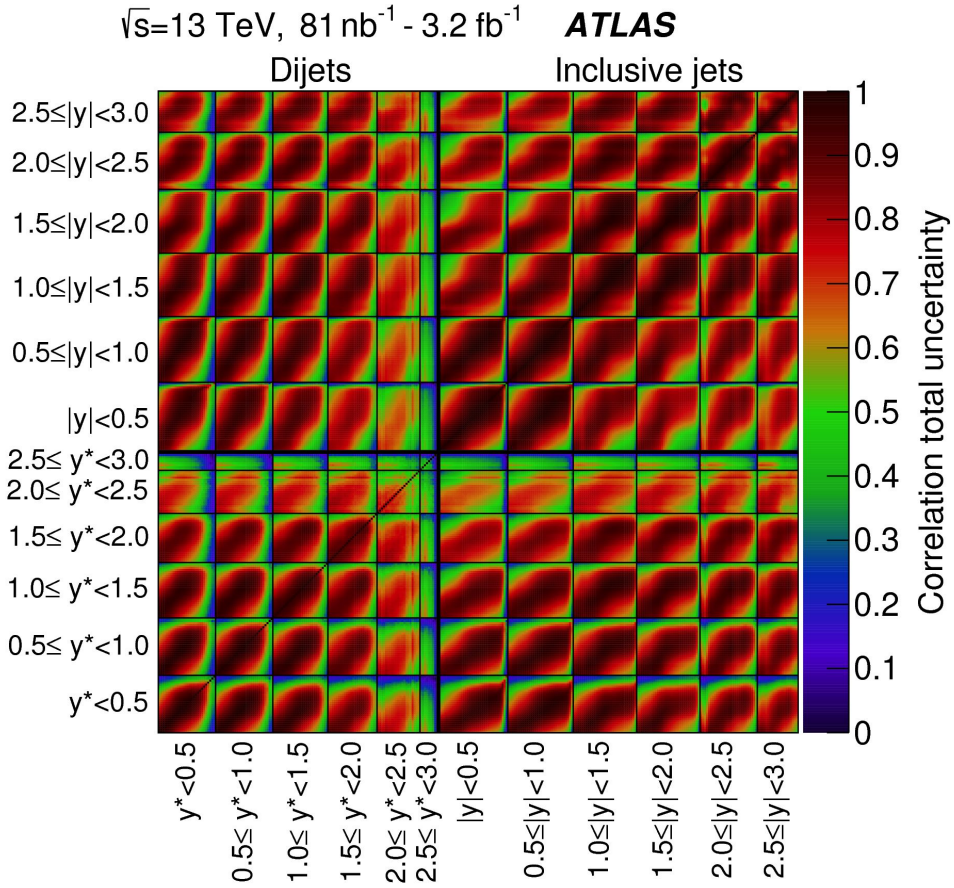
Propagation of systematic uncertainties from inputs

- Split of systematics in sub-components (fully correlated in phase-space, independent between each-other) allows to evaluate correlations between different phase-space regions and between different measurements

→ Relevant when effectively merging uncertainty components in ML-based methods

- Information made available in HEPData tables (http://hepdata.cedar.ac.uk/)

$$Cov_{ij} = \sum_{k=1}^{N_{syst}} s_i^k \cdot s_j^k$$

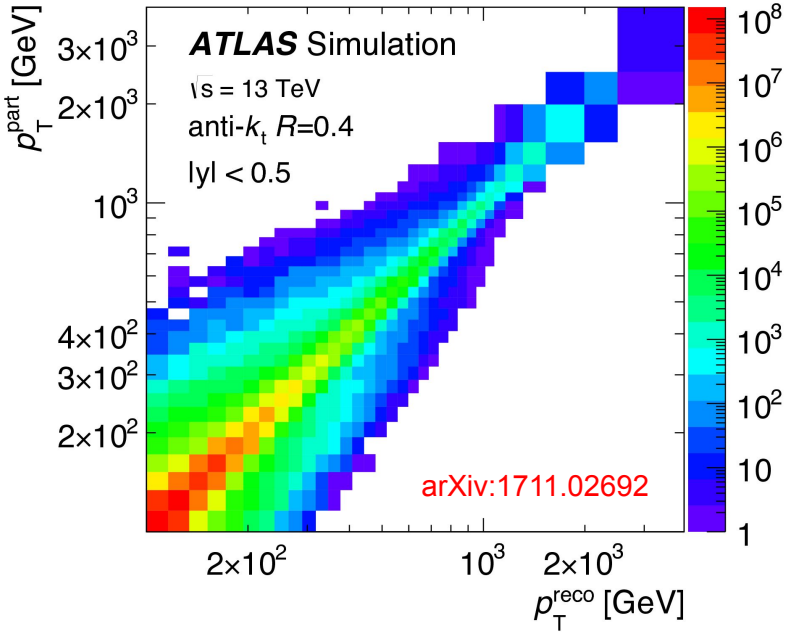


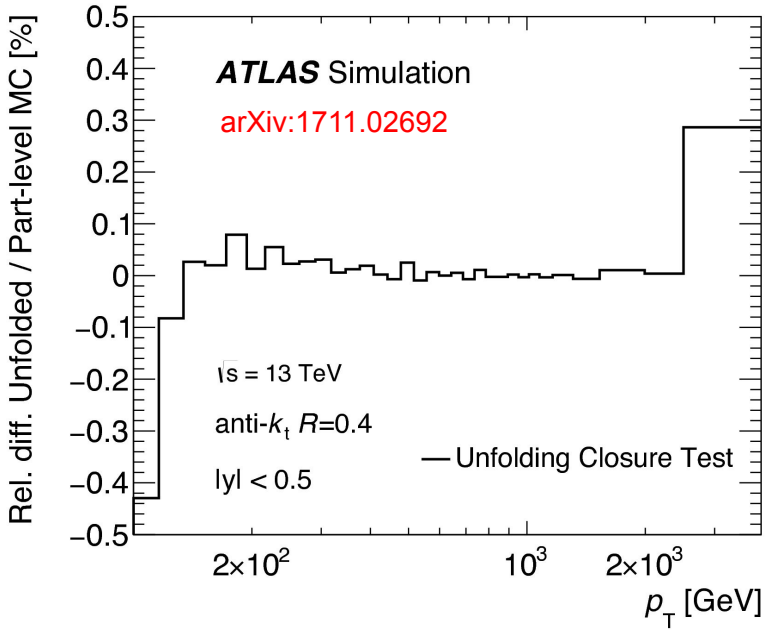
arXiv:1711.02692

Back

Data-driven closure test: motivation, procedure, example

- In-situ (i.e. *realistic*) determination of the unfolding uncertainty related to the data/MC shape difference and to the regularization :
  - reweight true MC by (smooth) function: improved data/recoMC agreement
    Reweighting performed within fine bins / event-by-event (arXiv:1711.02692, arXiv:2404.06204)
  - unfold the reweighted reconstructed MC
  - compare with reweighted true MC





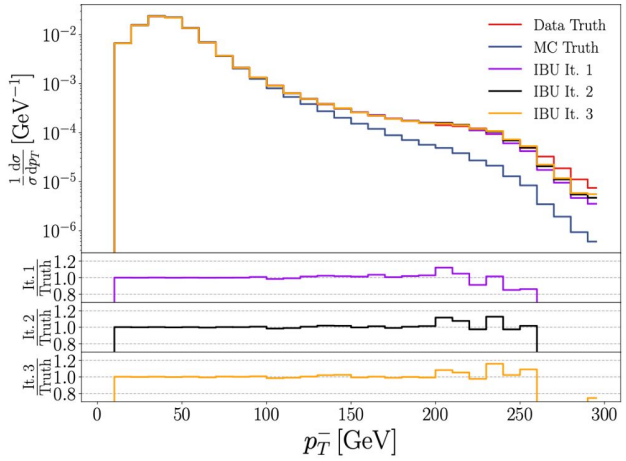
→ Applicable in cases without very different degenerated solutions (see eigenvalues of folding matrix, quality of the data/reweighted MC etc.)
In other cases allows to learn about the ill-posedness of the problem

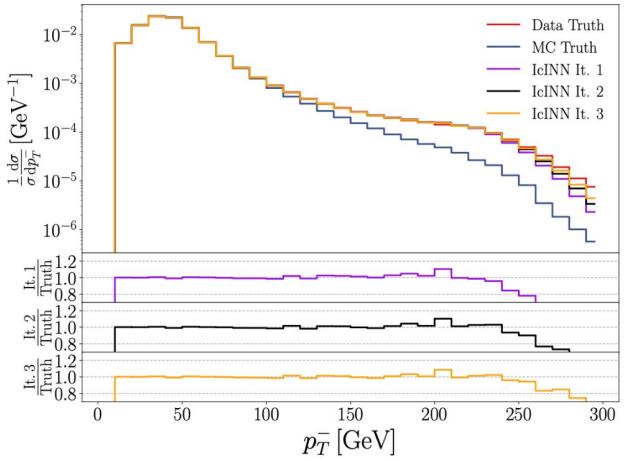
→ Method introduced in arXiv:0907.3791, used in arXiv:1112.6297 etc. … arXiv:2405.20041 (Omnifold 24-d) …

Back

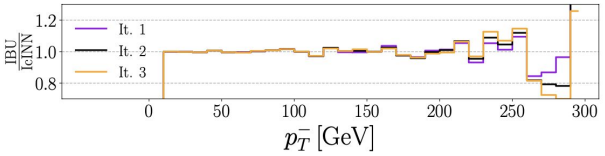
Comparison of Transfer Matrix- / ML-based unfolding

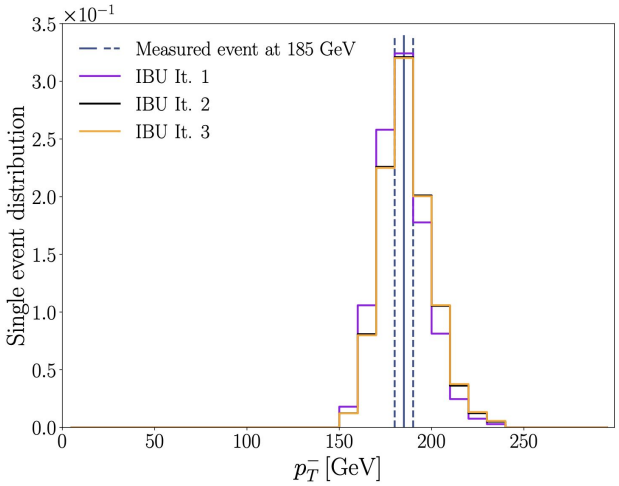
- Detailed comparison challenging: typically performed for the full unfolded distributions
- *New*: even-by-event comparison arXiv:2310.17037

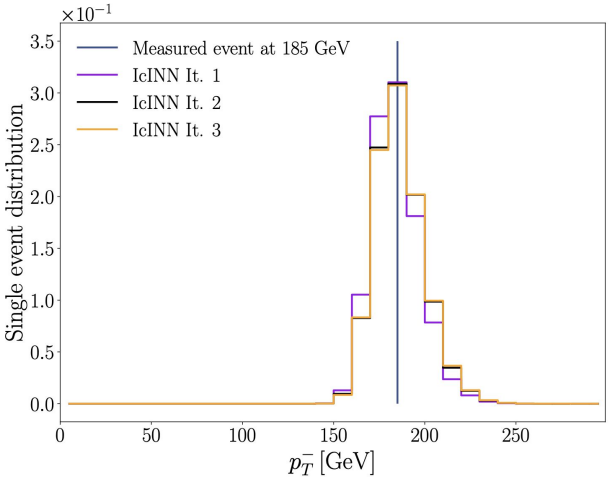


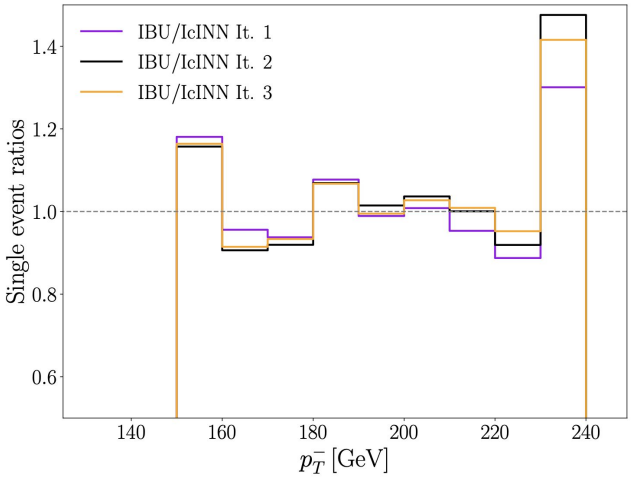


$pp \to Z\gamma\gamma,\ Z \to \mu^-\mu^+$









→ Challenging to visualise multidimensional results & uncertainties & correlations & phase-space coverage & biases & …

Back

Instead of Conclusions

- Numerous topics on which we can have interesting discussions

Thank you !!!